Title: **Maximum Likelihood Fitting of the Space-Time Autoregression with Exogenous Variables Model for Forecasting Groundwater Heads**

Author: Hosung Ahn, Hydrogeology Division, WRE, South Florida Water Management District

Target Journal: Water Resources Research

Forecasting future groundwater levels is important to water managers, who are concerned with the management of limited water resources during drought periods. Since historical monthly groundwater flows are highly dependent in space and time, the stochastic time series model of groundwater flow should incorporate a structure of corresponding temporal and spatial correlation. However, no such space-time stochastic model currently exists for use in south Florida. Thus, the objective of this paper is to report on the development of a space-time stochastic time series model and to show its usefulness in south Florida.

A new form of time series model (STARX) is presented. The new model allows time series information to be considered with the spatial structure of the system being modeled. A new parameter estimation algorithm is also proposed, which enables the modeler to accelerate convergence of parameter calibration and reduce both computational time and memory space significantly. The proposed model using this estimation algorithm has been applied to a multi-layered groundwater system in Collier County, Florida.

Results of model verification showed that the STARX model using the estimation algorithm not only preserves the historical statistics of mean, standard deviation, and space-time correlations, but is also ideal for forecasting large scale problems. The primary purpose of this paper is to discuss the development of this space-time model with a parameter calibration procedure. The Collier County case shows a successful local application. The STARX model may have potential applications in forecasting regional groundwater levels in the other regions, and solving general scientific and engineering problems involving high space-time correlation coefficients.

Title: **Maximum Likelihood Fitting of the Space-Time Autoregression with Exogenous Variables Model for Forecasting Groundwater Heads**

December, 1994

By: **Hosung Ahn**, Ph.D., P.E.

Hydrogeology Division, WRE,

South Florida Water Management District

West Palm Beach, Florida

Abstract

To forecast spatially correlated monthly groundwater heads, a new form of space-time autoregression with exogenous variables (STARX) model is presented, which incorporates theories of multi-variate time series model and spatial structure of system. As a maximum likelihood fitting of the STARX model, the expectation-maximization with a simplified smoother estimator (EMSSE) algorithm is proposed, which uses the time-invariant concept with an assumption of complete (no-missing) data. The EMSSE algorithm, compare to the time-variant version, enables not only to accelerate convergence of parameter calibration, but to reduce both computational time and memory space significantly. The proposed model with estimation algorithm is applied to a multi-layered groundwater system in Collier County, Florida. Results of model verification showed that the fitted STARX model preserves the first four moments as well as space-time correlations of the historical data, and that the STARX model with the proposed estimation algorithm is ideally suitable for the large scale forecasting problems.

1. Introduction

Forecasting future groundwater heads is an important issue to water managers, who are concerned with management of limited water resources during drought periods. There exist a variety of stochastic forecasting techniques. One promising option is the Kalman filter forecasting with an underlying system model, which might be either deterministic or stochastic. The deterministic system model is physically-based and accurate, but complication arises when applied with the Kalman filter algorithm. Another problem in this approach is that it should accompany the forecasting of inputs. However, some input variables, like rainfall to the groundwater system, are so uncertain that forecasting them results in significant errors which make forecasting itself virtually meaningless. That is why the Kalman filtering with a stochastic time series model is commonly used in practical forecasting problems, and by the same reason, the Kalman filtering with a deterministic system model was eliminated from the consideration in this paper.

When a stochastic time series model is fitted to the system, identification of model structure and calibration of model parameters may become the main concerns. If spatial correlation is dominant as in the case of regional monthly groundwater heads, either multivariate autoregressive moving average (MARMA) models or space-time ARMA (STARMA) model is commonly applied. However, the pure MARMA model does not properly account for the spatial structure of the system or so called the cross-boundary effects [*Bennett*, 1979, page 8]. Thus, without properly discriminating neighbor stations, the pure MARMA model applied to large systems may spread its regression weights to the entire system due to a large degree of freedom. Also, the mathematics involved with the

sophisticated estimation procedures of the large scale MARMA model become burdensome. That is why a family of univariate ARMA models have been extensively used in practical groundwater forecasting problems [*Shih et al*, 1992; *Graham and Tankersley*, 1993].

A family of STARMA models offers a way of generalizing both the ARMA models and the simultaneously specified spatial model by a hierarchical spatial ordering of neighbors. *Cressie* [1991, page 449] summarized several existing STARMA models, among them the following two models were of interest since their structures are similar to that of the proposed model. The STARMA model considered in *Pfeifer and Deutsch* [1980] is of the form

$$x_t = \sum_{k=1}^{Np} \sum_{i=0}^{\lambda_k} \phi_{ki} W^{(i)} x_{t-k} + w_t - \sum_{k=1}^{Nq} \sum_{i=0}^{m_k} \theta_{ki} W^{(i)} w_{t-k} \qquad (1)$$

where for the nx fixed spatial locations, x_t is the (nx×1) state vector at time t, Np is the temporal autoregressive order, Nq is the temporal moving average order, λ_k and m_k are the k-th spatial orders of autoregressive and moving average terms, ϕ_{ki} and θ_{ki} are the parameters (scalar) at k-th temporal order and i-th spatial order, respectively, $W^{(i)}$ is the (nx×nx) matrix of weight for the i-th spatial order, and w_t is the white noise vector. The other form of STARMA model with exogenous term given by *Stoffer* [1985] is defined by

$$x_t = \sum_{i=1}^{Nq} \Lambda_i D_i x_{t-i} + \sum_{i=0}^{Nk} \Psi_i z_{t-i} + w_t \qquad (2)$$

where, with the nx fixed stations, Λ_i is the (nx×nx) diagonal space-time transition intensity matrix at the i-th temporal lag, D_i is a known (nx×nx) distance matrix which expresses the spatial relationship in the random field x_t, z_t is the (nz×1) covariate vector having nz fixed

locations in a covariate space, and Ψ_i is the (nx×nz) regression matrix.

In the above two models, the spatial parameters are pre-determined by either the inverse distance weighing scheme or spatial statistics of covariance or semi-variogram. Then, parameter calibration is used to find the optimal time dependent multipliers to the predefined spatial structures. A difficulty in applying the above approaches is that it is hard to imagine how spatial dependence can arise other sources than integration of causation over the system [*Cressie*, 1991, page 450]. That is, defining the spatial weights $W^{(i)}$ or D_i explicitly, rather than from space-time correlation structure itself, may circumvent the real correlation system. Also, such lumped spatial parameters over the entire space will lose the integrity of the local correlation structures of the true system.

To overcome the above problems, this paper proposes a new form of the Space-Time Auto-Regression with Exogenous variables (STARX) model suitable for a large scale groundwater head forecasting problems. The large scale system referred in here is a system which has more than 20 state variables, enough to cause divergence problem occasionally during parameter calibration. Also included is a parameter calibration procedure by the maximum likelihood (ML) method, as well as an application to the multi-layered groundwater system in Collier County, Florida. The results of application revealed that the proposed estimation algorithm provides an adequate tool for the simultaneous estimation of both model parameters and error covariance terms, and that the fitted STARX model preserves the first-four moments as well as space-time correlations of the historical data.

2. The Proposed STARX Model

The proposed STARX model is based on the theory of MARMA model with incorporation of the spatial structure via the concept of spatial neighborhood. It is obvious that adding exogenous term (covariate) improves estimation of state variable (variate) and becomes a more physically-based model. Also use of covariate term, instead of the simultaneous modeling of both variate and covariate, enables to reduce the model dimension which is very critical in the calibration of a large scale system. The reasons for eliminating the moving average term in the proposed model are that the moving average term causes severe non-linearity that often adds difficulties in parameter calibration.

To formulate the STARX model, suppose that a spatial random vector denoted by x_t at time t, with t=1,...,T, is of interest to an investigator. With the nx fixed locations in space, the decomposition of x_t is $x_t'=(x_{t,1},...,x_{t,nx})$ where (′) indicates the transpose of either vector or matrix. Further it is supposed that a spatial random vector of covariate denoted by $z_t' = (z_{t,1},...,z_{t,nz})$ at time t may be measured concurrently from the nz fixed locations. For a regional groundwater head forecasting problem, $x_{t,j}$ may represent a potentiometric head at time t and site j in a given aquifer, whereas $z_{t,j}$ may be either rainfall, evapotranspiration, temperature, adjacent layer's head, or a composite of them, at time t and site j. Then, the STARX model which describes the current state x_t in terms of the previous states $\{x_{t-1}, ..., x_{t-Nq}\}$ and the covariate $\{z_t, z_{t-1}, ..., z_{t-Nk}\}$ can be expressed by

$$x_t = \sum_{i=1}^{Nq} D_i \circ \Lambda_i \, x_{t-i} + \sum_{j=0}^{Nk} E_j \circ \Omega_j \, z_{t-j} + w_t \qquad (3)$$

where Nq and Nk are the **temporal orders** of regression for x_t and z_t, respectively, Λ_i is the (nx×nx) matrix of parameters for the vector x_{t-i}, Ω_j is the (nx×nz) matrix of parameters for the covariate z_{t-j}, D_i is the known (nx×nx) **spatial index matrix (SIM)** for the x_{t-i}, E_j is the known (nx×nz) SIM for the z_{t-j}, and w_t is the (nx×1) multi-Gaussian white noise vector having a covariance of Q. Either D_i or E_j needs not be a symmetric matrix. Notation (∘) is the Hadamard product which is an element-wise product of two matrices of the same size [*Horn and Johnson*, 1985, page 321]. It is assumed that the Hadamard product has higher precedence than that of matrix multiplication. Then, the relationships of both $A \circ BC = (A \circ B)C$ and $A \circ BC \neq A \circ (BC)$ are satisfied, where $\{A, B, \text{and } C\} \in M_{m,n}$, with $M_{m,n}$ is a (m×n) vector space.

The m-th row and n-th column element $d_{i,mn}$ in D_i matrix is defined by

$$d_{i,mn} = \begin{cases} 1 & \textit{if } m\textit{- and } n\textit{-th stations are an } i\textit{-th time lag neighbor} \\ 0 & \textit{otherwise} \end{cases}$$

with a case of m=n as a default neighbor. The same definition is applied to the element $e_{j,mn}$ in E_j matrix, indicating that $x_{t,m}$ and $z_{t,n}$ are the j-th time lag neighbor. The application

where the dimensions of the matrices in (4) are as follows: D(nx×ns), E(nx×nc), Λ(ns×ns), Ω(nx×nc), x(t-1)(ns×1), z(t)(nc×1), $\Phi=[\Phi_1,...,\Phi_{Nq}]$ (nx×ns), and $\Psi=[\Psi_0,...,\Psi_{Nk}]$(nx×nc), with ns=nx×Nq, and nc=nz×(Nk+1). The third expression in (4) is exactly the same format as that of a multivariate autoregressive model with a covariate term. However, it should be noted that both Φ and Ψ in (4) are sparse matrices due to the Hadamard products. With properly determined temporal orders Nq and Nk by the model identification procedure, the equation (3) or (4) was defined as a STARX model of order (Nq,Nk), or simply a **STARX(Nq,Nk) process.**

The proposed STARX model is different from the approaches in (1) or (2) since it imposes the spatial structure of the system via SIMs without pre-constraining its parameters. The STARX model is more flexible in its format than a MARMA model. That is, if the D_i's in Equation (3) are diagonal matrices with no covariate, the STARX model becomes a set of univariate autoregressive models. Also, if the D_i's and E_i's are unit matrices, the STARX model is nothing but a multivariate autoregressive model with an exogenous term.

where Nq and Nk are the **temporal orders** of regression for x_t and z_t, respectively, Λ_i is the (nx×nx) matrix of parameters for the vector x_{t-i}, Ω_j is the (nx×nz) matrix of parameters for the covariate z_{t-j}, D_i is the known (nx×nx) **spatial index matrix (SIM)** for the x_{t-i}, E_j is the known (nx×nz) SIM for the z_{t-j}, and w_t is the (nx×1) multi-Gaussian white noise vector having a covariance of Q. Either D_i or E_j needs not be a symmetric matrix. Notation (∘) is the Hadamard product which is an element-wise product of two matrices of the same size [*Horn and Johnson*, 1985, page 321]. It is assumed that the Hadamard product has higher precedence than that of matrix multiplication. Then, the relationships of both A∘BC=(A∘B)C and A∘BC≠A∘(BC) are satisfied, where {A, B, and C}$\in M_{m,n}$, with $M_{m,n}$ is a (m×n) vector space.

The m-th row and n-th column element $d_{i,mn}$ in D_i matrix is defined by

$$d_{i,mn} = \begin{cases} 1 & \textit{if } m\textit{- and } n\textit{-th stations are an } i\textit{-th time lag neighbor} \\ 0 & \textit{otherwise} \end{cases}$$

with a case of m=n as a default neighbor. The same definition is applied to the element $e_{j,mn}$ in E_j matrix, indicating that $x_{t,m}$ and $z_{t-j,n}$ are the j-th time lag neighbor. The application section will show an example of how to determine SIMs by the first spatial order neighbors using the Thiessen polygons.

The STARX model in (3) can now be rewritten in the block matrix forms of

$$\begin{aligned} x_t &= [D_1 \ \cdots \ D_{Nq}] \circ [\Lambda_1 \ \cdots \ \Lambda_{Nq}] \begin{bmatrix} x_{t-1} \\ \vdots \\ x_{t-Nq} \end{bmatrix} + [E_0 \ \cdots \ E_{Nk}] \circ [\Omega_0 \ \cdots \ \Omega_{Nk}] \begin{bmatrix} z_t \\ \vdots \\ z_{t-Nk} \end{bmatrix} + w_t \\ &= D \circ \Lambda x(t-1) + E \circ \Omega z(t) + w_t \\ &= \Phi x(t-1) + \Psi z(t) + w_t \end{aligned} \tag{4}$$

where the dimensions of the matrices in (4) are as follows: D(nx×ns), E(nx×nc), Λ(ns×ns), Ω(nx×nc), x(t-1)(ns×1), z(t)(nc×1), $\Phi=[\Phi_1,...,\Phi_{Nq}]$ (nx×ns), and $\Psi=[\Psi_0,...,\Psi_{Nk}]$(nx×nc); with ns=nx×Nq, and nc=nz×(Nk+1). The third expression in (4) is exactly the same format as that of a multivariate autoregressive model with a covariate term. However, it should be noted that both Φ and Ψ in (4) are sparse matrices due to the Hadamard products. With properly determined temporal orders Nq and Nk by the model identification procedure, the equation (3) or (4) was defined as a STARX model of order (Nq,Nk), or simply a **STARX(Nq,Nk) process.**

The proposed STARX model is different from the approaches in (1) or (2) since it imposes the spatial structure of the system via SIMs without pre-constraining its parameters. The STARX model is more flexible in its format than a MARMA model. That is, if the D_i's in Equation (3) are diagonal matrices with no covariate, the STARX model becomes a set of univariate autoregressive models. Also, if the D_i's and E_i's are unit matrices, the STARX model is nothing but a multivariate autoregressive model with an exogenous term.

3. ML Fitting by the Expectation-Maximization Algorithm

There are a several different parameter estimation techniques available for the stochastic time series models. These methods includes the method of moment, least square method, ML method, etc. The ML method attempts to incorporate all the information into a model by working with the complete distribution of the measurement. This is why ML estimates are preferred over the other methods for stochastic time series models. If the objective is just fitting of the STARX model, one may use the Gauss-Newton algorithm or the

scoring algorithm [*Harvey*, 1990, page 134]. However, the ultimate objective of fitting the proposed STARX model in here is forecasting by the Kalman filter algorithm. The Kalman filter seeks to provide the minimum error variance estimator for the state vector with balancing both the model and output errors, given the measurements. Thus, the Kalman filter forecasting requires system model parameters, noise covariance terms, and the updated state and error covariance terms. Those of which should be calibrated simultaneously, requiring a special algorithm rather than conventional optimization methods.

It was known that the EM algorithm [*Dempster et al.*, 1977] can be applied in conjunction with the modified Kalman smoother estimators to derive a simple recursive procedure for the ML fitting of multivariate ARMA models [*Stoffer*, 1985; *Shumway*, 1988, p. 173]. The EM algorithm is an alternative non-linear optimization algorithm ideally suitable for an incomplete data set. The term "incomplete data" implies the existence of two sample spaces: a measured part and an unmeasured part. With some incompletely measured data Y which is also a function of a known signal process X and a noise process V, it was defined that $\ln L(X,V \mid \theta)$ is a log likelihood based on the complete data, and that $\ln L(Y \mid \theta)$ is a log likelihood based on the incomplete data, where θ is the parameter set of the model to be estimated. Both X and V are not directly measured, but only indirectly expressed through Y. Then, the EM algorithm is designed to find θ iteratively by maximizing the expectation of the complete-data log likelihood conditioned on the measured data Y. A conditional expectation of the log likelihood at the i-th iteration can be defined by

$$Q(\theta \mid \theta_i) = E_i \left[\ln L(X,V;\theta) \mid Y;\theta_i \right]. \tag{5}$$

Then, the expectation step (E-step) computes $Q(\theta \mid \theta_i)$ from the above log likelihood function of the time series model, and the maximization step (M-step) chooses θ_{i+1} to maximize $Q(\theta \mid \theta_i)$ using one of the optimization techniques. Since the X process cannot be measured directly, the above equation (5) can be written in terms of the Kalman smoother estimators.

To apply the Kalman filter recursion, the STARX model should be transformed into the state-space form, which consists of the state and measurement equations. The **state equation** of the STARX model in (4) can be written by

$$x(t) = \Phi x(t-1) + \Psi z(t) + w(t) = [\Phi \ \Psi]\begin{bmatrix} x(t-1) \\ z(t) \end{bmatrix} + w(t) = \theta X(t-1) + w(t) \qquad (6)$$

where $x(t)'=[x_t', \ldots, x_{t-Nq+1}']$(ns×1), $z(t)'=[z_t', \ldots, z_{t-Nk}']$(nc×1), $w(t)'=[w_t', 0, \ldots, 0]$(ns×1) are augmented vectors, and the parameter matrices are given by

$$\Phi = \begin{bmatrix} \Phi_1 & \cdots & \cdots & \Phi_{Nq} \\ I & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix}, \quad \textit{and} \quad \Psi = \begin{bmatrix} \Psi_0 & \cdots & \Psi_{Nk} \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

with **I** and **0** are the identity and zero matrices. To allow for the possibility of missing data and the existence of measurement noise, the **measurement equation** is written by

$$y_t = M(t)\, x(t) + v_t \qquad (7)$$

where y_t is the (nx×1) measurement vector at time t, $M(t)=[M_t,0,\ldots,0]$ is the (nx×ns) measurement matrix (whose element $\{m_{ij}\}$ in M_t is 1 if i=j, or 0 if i≠j), and the **measurement noise** v_t is the (nx×1) multi-Gaussian white noise having $v_t \approx N(0,R)$. With (6) and (7), the joint log likelihood of the complete data x(1),...,x(T) [*Stoffer*, 1985] is

$$\begin{aligned} \ln L \approx & -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}[x(0)-\mu]' \Sigma^{-1}[x(0)-\mu] \\ & -\frac{T}{2}\ln|Q^*| - \frac{1}{2}\sum_{t=1}^{T}[x(t)-\theta X(t-1)]' Q^{*-1}[x(t)-\theta X(t-1)] \\ & -\frac{T}{2}\ln|R| - \frac{1}{2}\sum_{t=1}^{T}[y_t-M_t x(t)]' R^{-1}[y_t-M_t x(t)] \end{aligned} \tag{8}$$

where $\mu=E[x(0)]$, $\Sigma=cov[x(0)]$, and $Q^*=E[w(t)w(t)']$.

The problem of estimating x(t) in (6) can be approached by the expectation of it conditioning on the measurements $y_1,...,y_T$ and $z_1, \ldots, z_T$ as

$$\hat{x}_t^s = E\,[x(t) \mid y_1,...,y_s, z_1,...,z_s] \tag{9}$$

where s is the span of the measurement. Defining that the estimation error $\tilde{x}_t$ is the true value x_t^* minus the estimated value $\hat{x}_t^s$, that is, $\tilde{x}_t=(x_t^* - \hat{x}_t^s)$, the error covariance is expressed by

$$p_t^s=E\,[(x_t^*-\hat{x}_t^s)(x_t^*-\hat{x}_t^s)' \mid y_1,...,y_s, z_1,...,z_s] \tag{10}$$

and the one-time lag error covariance may be written by

$$p_{t,t-1}^s=E\,[(x_t^*-\hat{x}_t^s)(x_{t-1}^*-\hat{x}_{t-1}^s)' \mid y_1,...,y_s, z_1,...,z_s]\ . \tag{11}$$

Then, the following three problems occur when estimating the $\hat{x}_t^s$, p_t^s, and $p_{t,t-1}^s$: if t=s, it is called the **filtering** problem; if t<s, it is the **smoothing** problem; and if t>s, it is the **forecasting** problem. From the above definitions of conditional expectations, the modified Kalman smoother is derived. The results applied to the STARX model with the corresponding the EM steps were summarized in Appendix.

The following is the summarized estimation procedure: With an assumed initial parameter set

$$\Xi(0) = \{\theta(0), Q^*(0), R(0), \mu(0), \textit{and } \Sigma(0)\}$$

where (0) indicates the initial step before iterations, the i-th iteration computes the Kalman smoother estimators and the E-step (A4 through A16) with computing the following incomplete-data log-likelihood [*Gupta and Mehra*, 1974; and *Shumway and Stoffer*, 1981]

$$-2\ln L(Y;\theta) \approx \sum_{t=1}^{T} \ln |M(t)\, p_t^{t-1} M(t)'| + \sum_{t=1}^{T} [y_t - M(t)\hat{x}_t^{t-1}]' [M(t)\, p_t^{t-1} M(t)' + R]^{-1} [y_t - M(t)\hat{x}_t^{t-1}] \quad (12)$$

where $\hat{x}_t^{t-1}$ and p_t^{t-1} are from the forward recursion. Then, the M-step recalibrates the parameters θ_{i+1}(A17), Q^*_{i+1}(A18), and R_{i+1}(A19). The reason why the ML function is multiplied by -2 is that this ML function is the same as that of the least square objective function. Then, the objective function of the optimization is minimizing equation (12).

4. Simplified Kalman Smoother for Complete-data Sets

The above EM algorithm with the Kalman smoother estimator, although it has a greatest advantage when it is used for incomplete-data sets, can be simplified when a complete-data set is available. The expectation-maximization with a simplified smoother estimator (EMSSE) was derived with time-invariant error covariance terms in both the forward and backward recursions. The concept of time-invariance in Kalman filtering was briefly discussed by *Lainiotis* [1978], where the partitioned filter was solved in the form of time-invariance and steady-state (Weiner) filter.

Without missing data, the measurement equation can be written by

$$y_t = M\, x(t) + v_t \tag{13}$$

where the measurement matrix (nx×ns) becomes M=[I,0,...,0], with I and 0 are identity and zero matrices, respectively, each of which has a dimension of (nx×nx). Since M is time-invariance, p_t^{t-1}, p_t^t, and K_t in the forward recursion are also time-invariance. Using $p_1(=p_t^{t-1})$, $p_0(=p_t^t)$, and $K(=K_t)$, equations (A5), (A6), and (A8) can be rewritten as

$$\left.\begin{cases} p_1 = \Phi p_0 \Phi' + Q & ...(a) \\ K = p_1 M' \,[M\, p_1 M' + R]^{-1} & ...(b) \\ p_0 = (I - KM)\, p_1 & ...(c) \end{cases}\right\} \tag{14}$$

which can be estimated independently from the state estimation in each iteration. The equation (14) is a set of simultaneous equations having three unknowns and three known terms (Φ, Q, and R). The explicit solution of (14) is unknown to the author, but an iteration method was used: Since the Kalman forward recursion is designed to obtain the minimum mean-squared error estimators, optimal p_1, K, and P_0 in each iteration can be obtained, with an initial assumption of p_0 which satisfies the positive semi-definition matrix, by solving the above three equations sequentially and repeatedly until the estimators are stabilized. The experiment showed that less than 10 iteration leads to a good convergence in general.

For the backward recursion, the time-invariant estimator $J(=J_t)$ is obtained by

$$J = p_0 \Phi' \,[p_1]^{-1} \tag{15}$$

and, with notations of $p^T{}_0(=p^T{}_t)$ and $p^T{}_1(=p^T{}_{t-1,t-2})$, the smoother error covariances are

$$p_0^T - J\, p_0^T J' = p_0 - J\, p_1 J' \tag{16}$$

$$p_1^T - J\,p_1^T J' = p_0 J' - J\phi\,p_0 J' \tag{17}$$

Again, since the above equations are derived under the condition of minimum mean-squared error estimation, an iteration of each equation, with an initial assumption of it which satisfies the positive semi-definition matrix, will result in the optimal p_0^T and p_1^T.

Alternatively, equations (16) and (17) can be solved explicitly by the Kronecker sum properties [*Bellman*, 1970, page 125]: Since both (16) and (17) have the form of

$$X - UXV = W \tag{18}$$

where X is an unknown matrix, and U, V, and W are known coefficient matrices, the solution of the above equation to X is given by

$$X = \sum_{k=0}^{\infty} U^k W V^k \,. \tag{19}$$

In practice, summation in the above equation may continue until the term U^kWV^k at k is insignificant. The condition in (19) is that the sum of any two characteristic roots (eigenvalues) of U and V is non-zero. This condition will be satisfied when both U and V are the stability matrices which satisfy that all of the eigenvalues of U and V have negative real parts. Since V=U′ for both (16) and (17), and J, p_0, and p_1 are all positive semi-definite matrices, the characteristic polynomials of both U and V are the same and the stability matrix conditions are automatically satisfied.

After solving the error covariances in equations (14) through (17), the forward recursion for t=1,...,T is given by

$$\hat{x}_t^{t-1} = \Phi \hat{x}_{t-1}^{t-1} + \Psi z(t) \tag{20}$$

$$\hat{x}_t^{t} = \hat{x}_{t-1} + K(y_t - M x_t^{t-1}) \tag{21}$$

with computing the log likelihood function in (12) with p_1 instead of p_t^{t-1}. Then, the backward recursion for t=T,T-1,...,1 is computed by

$$\hat{x}_{t-1}^{T} = \hat{x}_{t-1}^{t-1} + J(\hat{x}_t^{T} - \Phi \hat{x}_t^{t-1}) \tag{22}$$

and the EM estimates are computed by (A15) through (A19) in Appendix.

The great advantage of EMSSE is that whenever complete-data is available, the computational load is significantly reduced due to the reduced number of matrix inversions for both K and J. The larger the system is, the greater is the reduction of computational loads. Also, the EMSSE algorithm eliminates requirement of a large memory space for p_t^t, which should have a dimension of $[(nx \times Nq)^2 T]$ to link the forward and backward recursions. The other advantage is that use of time-invariant parameters eliminates the effect of the initial assumption of the error covariances (p_0, p_1, p_0^T, and p_1^T) provided in each iteration of estimation, and consequently escalates the convergence rate of the estimation algorithm.

5. Application

5.1. Description of the Model Area and Historical Groundwater Head Data

The model area consists of the western half of Collier County and a small portion of the southwestern Lee County, Florida. This area is characterized by moderately drained sandy soils with extensive agricultural and urban development. This area receives an average of 1380 millimeters of rainfall annually, with about 60 percent occurring during the wet

season (mid May-October). Without having a significant surface water inflow source or storage, the area relies mainly on the groundwater sources to meet water use demands. Thus, this area typically experiences water shortage problems. The area's hydrogeology consists of four underlying fresh water aquifers as described in Table 1, which provide most of the region's water supplies. The top three aquifers are separated from each other by semi-confining beds having low permeable materials, however a significant amount of inter-aquifer flows take place due to the leaky structure. These three aquifers are extensively developed because of their low salinity content. The mid-Hawthorn aquifer is separated from the sandstone aquifer by a low permeable confining unit, and not highly utilized due to increased well depth, high salinity, and low yield. Currently, the U.S. Geological Survey (USGS) maintains about 150 monitoring wells in this area, from which piezometric heads are collected. After eliminating stations whose records are inconsistent or having sampling intervals greater than a month, only 115 gaging stations were selected (refer to Table 1 and Figure 1). The periods of record range from early 1900's through August 1993, and a full-fledged monitoring program began in 1987; from that time, a complete-data is available.

A volumetric water budget analysis in this area by *Bennett* [1992] showed that layer 1 receives 95% of its inflow from the local rainfall and releases about 16% of its outflow to layer 2. Layer 2 receives about 82% of its inflow from the layer-1 and releases only 9% of its outflow to layer 3. Layer 3 receives 54% of its inflow from layer 2 and 43% of it from layer 4 while releasing 40% of outflow to layer 2. Layer 4 receives most of its inflow through upward leakage from the deeper aquifer and releases most of its outflow to layer 3. These results indicates that layer 1 and 4 are not significantly affected by the adjacent layers,

while layer 2 is influenced by layer 1, and layer 3 is affected by both layer 2 and 4.

Experiment of the EMSSE algorithm in this study revealed that a STARX model having more than 50 state variables was extremely inefficient in calibration. Thus, instead of a STARX model comprising all aquifer layers simultaneously, the following three separate STARX models were developed:

Model I : for layer 1 with no-covariate,

Model II : for layer 2, with layer 1 as covariate, and

Model III : for layer 3 and 4, with layer 2 as covariate.

5.2. Treatments of the Raw Data

Seasonality

Seasonality in time series is a regular change in the data values that occurs at the same time in a given period. The monthly groundwater heads in the model area have strong seasonalities (Figure 2). Unlike monthly means, the monthly standard deviations during the wet season are smaller than those of the dry season except for the Surficial aquifer, because most summer rainfalls induce near maximum groundwater heads due to the high permeable confining units. Several methods are available to handle the seasonality in time series model, among which the multiplicative explanatory variable approach by the standardization procedure [*Salas et* al., 1980, page 241] was adopted, because this approach is very simple in nature and does not increase the number of model parameters. If a time series $u_{t,i}$ (untransformed) at a site i is measured during t=1,...,T, $u_{t,i}$ can be rewritten to $u_{\tau,\upsilon,i}$ using the seasonal notation $\tau(1,..,\omega)$ and year $\upsilon(=1,...,N)$, where ϖ is a number of season (12 month) in

a year and N is the number of years, with T=ϖN. Then the standardization

$$x_{\tau,v,i} = \frac{u_{\tau,v,i} - \mu_{\tau,i}}{\sigma_{\tau,i}} \tag{23}$$

will remove the seasonality, where $\mu_{\tau,i}$ and $\sigma_{\tau,i}$ are the seasonal mean and standard deviation, respectively, at site i and month τ. After standardization, the series $x_{\tau,v,i}$ should be written as $x_{t,i}$ series to fit for the STARX model. This standardized series $x_{t,i}$ has a mean of 0 and a variance of 1, if each month's data meets the normality condition. Then, the correlation coefficient of this standardized series is the same as covariance.

Normality

Several statistical tests are available for testing the hypothesis that a given series is normally distributed. The skewness test of normality [*Salas et al.*, 1985, page 92] is based on the fact that the skewness coefficient for a normal variable is zero. An estimation of the skewness of a time series $u_{t,i}$, t=1,...,T is given by

$$\gamma_i = \frac{\frac{1}{T}\sum_{t=1}^{T}(u_{t,i}-\mu_i)^3}{\left[\frac{1}{T}\sum_{t=1}^{T}(u_{t,i}-\mu_i)^2\right]^{3/2}} \tag{24}$$

where μ_i is the sample mean at site i. The 1-α probability limits is defined by $\pm u_{1-\alpha/2}(6/T)^{1/2}$, where $u_{1-\alpha/2}$ is the 1-α/2 quantile of the standard normal distribution. If the estimated γ_i is within the probability limit, the hypothesis of normality is accepted, or otherwise rejected. If data is non-normal, several transformation techniques are available. Since some groundwater heads in the model area did not satisfy the hypothesis of normality, the following Box-Cox

transformations [*Bras and Rodriguez-Iturbe*, 1985, page 73] was used:

$$x_{t,i} = \begin{cases} \dfrac{(u_{t,i})^{\lambda_i}-1}{\lambda_i} & \lambda_i \neq 0 \\ \ln(u_{t,i}) & \lambda_i = 0 \end{cases} \qquad (25)$$

where λ_i is the transformation parameter to be estimated. Using a simple trial and error method, λ_i's, i=1,...115, were calibrated, whose values range from -0.6 to 2.2 with an average of 0.43. After the Box-Cox transformation, the γ_i's (mean ± standard deviation of γ_i's) were improved from -0.065±0.35 to -0.042±0.26. Moreover, the number of sites accepted by the normality assumption with α=0.1 was increased from 92 to 106 sites (92.2% accepted). In addition, the logarithm, exponential, and power transformations were tried to improve normality conditions, but the Box-Cox transformation was superior to the other methods.

Temporal Trends

One of the common non-stationarities in groundwater head data was a temporal trend which is a long, consistent change in the time series values from beginning to end. Temporal trends can be modeled by either linear, quadratic, or cubic pattern. A simple linear trend model [*Shumway*, 1988, page 124] was used which is the form of

$$u_{t,i} = \beta_{0,i} + \beta_{1,i}\, t + e_{t,i} \qquad (26)$$

where $e_{t,i}$ is the model noise, $\beta_{0,i}$ and $\beta_{1,i}$ are regression parameters estimated by

$$\beta_{0,i} = u_i^* - \beta_{1,i}\, t^* \qquad (28)$$

$$\beta_{1,i} = \frac{\sum_{t=1}^{T} (t-t^*)(u_{t,i}-u_i^*)}{\sum_{t=1}^{T} (t-t^*)^2} \qquad (27)$$

where t^* and u_i^* are the means over the time axis and observed sample series, respectively. Then, the detrended series can be modeled by

$$\begin{aligned} x_{t,i} &= u_{t,i} - \beta_{0,i} - \beta_{1,i}\, t \\ &= (u_{t,i} - u_i^*) - \beta_{1,i}\,(t-t^*)\ . \end{aligned} \qquad (29)$$

After standardization and the Box-Cox transformation, the temporal trend in each site was removed before parameter calibration. For i=1,...,115, the estimated $\beta_{0,i}$'s have a mean of -0.0465, a standard deviation of 0.144, and a maximum of 0.318; whereas the estimated $\beta_{1,i}$'s have a mean of 0.0036, a standard deviation of 0.009, and a maximum of 0.028.

5.3 Neighbor Sites Determined by the Thiessen Polygons.

As mentioned before, the SIM's matrices are constructed by the binary weights with the concept of neighbor which can be determined by the sampling grid. Sampling grid can be either regular or irregular spacings. For an irregularly spaced system, either the spatial statistics or the spatial ordering scheme can be used to determine the neighbors. For example of the spatial statistics, *Stoffer* [1986] used semi-variogram after assuming that the first and second spatial orders are stationary. If a fitted semi-variogram has a significant sill value, the range of the semi-variogram can be used as a boundary for selecting neighbors. However, it is usually difficult to find such a distinct sill or range in practical cases. Moreover, an assumption of spatial stationarity makes the procedure simple, but loses the local correlations

due to the use of the lumped spatial function.

The spatial ordering scheme is determined by the lagging of adjacency of boundary cells [*Bennett*, 1979 p. 481]. The first order sites are the ones which have right adjoined common boundaries to the site of interest. Spatial orders can be extended to the second or higher orders. However, *Bennett* [1979 p.484] pointed out that "extension to high-order lags becomes very dubious since there is no simple or objective rule by which the contiguity (lagged) counts can be made." Thus, the first spatial order was used in here to build the SIM matrices. Furthermore, the assumptions of D_1=,...,D_{Nq}=D and E_0=,...,E_{Nk}=E were used in the STARX model for simplicity in the STARX model.

Without having any natural geographic boundaries, the Thiessen polygons were used to define the boundaries of gaging stations. An advantage of the Thiessen polygons method in selecting neighbors is that it accounts for both distance and directional components simultaneously. The Thiessen polygon is a convex polygon whose boundary defines the area that is closest to the site of interest relative to all other sites [*Environmental System Research Institute*, 1992]. It is geometrically defined by the perpendicular bisectors of the lines which connect all neighbor sites (Figure 3a). When the covariate z_t is used in the STARX model, the covariate Thiessen polygons should be constructed site by site. That is, if the nx state vector and the nz covariate vector are used, a total nx maps should be prepared, each of which is constructed on the covariate space with adding a site from the variate space. Figure 3b shows examples of some covariate polygons. From those Thiessen polygons, all sites whose boundaries are joined by lines were considered as the first spatial order neighbors, and whose the corresponding elements in the SIM's are coded as 1, or otherwise 0. If two sites

were joined by a point (for example, site #7 and #12, or #26 and #40 in Figure 3a), they were classified as the second-order neighbors, but if two sites were joined by a point and a line (site #26 and #40 in Figure 3a), they were classified as the first-order neighbors.

5.4. Calibration of the STARX Model

To determine the temporal order of a multivariate autoregressive model, either the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) can be used. Using the residual sum of products $RSP=\Sigma^{T}_{t=1}[w_t' w_t]$, where w_t is the model noise estimated by the ML method, the form of AIC [*Shumway*, 1988, page 167] is given by

$$AIC(Nq)=\ln\left|\frac{RSP}{T}\right|+\frac{2\ nx^2\ Nq}{T} \quad . \tag{30}$$

where nx is the dimension of state vector, and T is the record length. Then the model order is selected whose AIC is smallest of all. The term ln|RSP|, which is the reduced likelihood [*Blockwell and Davis*, 1987, page 280], is equivalent to the first term of the ML function in (12). Note that if the conditional fitting procedure leading to the RSP is applied, T in equation (30) should be replaced by the number of effective observations (T-Nq). Since (nx^2Nq) is the number of effective parameters of an autoregressive model, the number of effective parameters for a STARX model is [$nx^2Nq+nx\times nz\times(Nk+1)$]. The BIC [*Shumway*, 1988, page 167] chooses the model order Nq that minimizes

$$BIC(Nq)=\ln\left|\frac{RSP}{T}\right|+\frac{nx^2Nq\ \ln T}{T} \quad . \tag{31}$$

Using the EMSSE algorithm, several alternative STARX models having different

temporal orders were calibrated, from which ML function, AIC, and BIC were computed (Table 2). The period of record of the calibration was from January 1978 to December 1992, while $\mu_{\tau,i}$ and $\sigma_{\tau,i}$ in (23) were computed by all available historical data, which range from 7 years to 42 years. Based on the estimated AIC's and the first term of ML function, STARX(2,no), STARX(1,0), and STARX(3,0) were selected for Model I, II, and III, respectively.

5.5. Verification

Two of the common applications of the stochastic time series model is stochastic data generation and forecasting. Thus, verification of the calibrated STARX model was focused on those two aspects.

Data Generation

When data generation, M_t in the measurement equation (7) is nothing but an identity matrix and can be ignored. Substituting all x_t terms in (3) with (y_t-v_t) and rearranging for y_t results in the following recursion equation:

$$y_t = \sum_{i=1}^{Nq} D_i \circ \Lambda_i y_{t-i} + \sum_{j=0}^{Nk} E_j \circ \Omega_j z_{t-j} + w_t + v_t - \sum_{i=1}^{Nq} D_i \circ \Lambda_i v_{t-i} \,. \tag{31}$$

With $x(0) \approx N(\mu,\Sigma)$, the initial condition of generated series becomes $y_0 \approx N(\mu,\Sigma+R)$. To generate data by (31), a complete set of covariate series z_t (which should also be generated beforehand), noise series w_t and v_t, and the assumed initial data y_0 should be provided.

Using the calibrated Model I, $y_{t,i}$ series (t=1,...,1600, i=1,...,48) were generated. After discarding the first several data (t=1,...,100) to eliminate the impact of initial measurement y_0, 5 sets of $y_{t,i}$ series (t=1,...,T with T=100, 300, 500, 1000, and 1500) were taken from the generated, and spatial averages of the first four moments were computed, some of which are scaled so that theoretical values of them are all zeros (Figure 4). Asymptotic values of them converge to near zeros, implying that the first four moments were preserved by the fitted STARX model. Also, an investigation of preserving the space-time correlations of the fitted STARX model was done using the generated time series (Figure 5). This comparison revealed that the correlations were reasonably well preserved, particularly those of zero time lag (cross-correlations).

Forecasting

By definition, forecasting is the unconditional prediction (ex ante prediction) [*Harvey*, 1991, page 14]. Since the EMSSE algorithm uses the state-space formulation of the STARX model, the Kalman forward recursion in the EMSSE algorithm can easily be applied to forecast the future state. That is, at the present time t, the **forecasted state** of the STARX model (4) for the lead time l can be given by

$$x(t+l) = \phi x(t-1+l) + \psi z(t+l) \qquad (32)$$

which is the minimum variance forecasting. The covariate z(t+l) should be forecasted beforehand. The forecasted error variance is simply given by

$$p_0(t+l) = \phi p_0(t-1+l)\phi' + Q^* \qquad (33)$$

where $Q^*=E[w(t)w(t)']$.

In order to show the forecasting ability of the three fitted STARX models, the groundwater heads for l=1,...,12 were forecasted at the end of December 1992. From the forecasting error covariance p_0 in equation (33), the spatial mean and standard deviation of the forecasted error ($p_0^{1/2}$) were computed (Table 3). The values in this table are those of standardized series having a mean of zero and a variance of one. The statistics at the first row (l<1) in this table are for the calibration case, which are constant during the period of calibration since complete data were used in each model. An interesting thing to note is that the forecasting errors increased significantly at the first lead time, then remained almost constant thereafter. Figure 6 shows the contour maps of both historical and forecasted groundwater heads for each layer at the selected lead times (l=1,2, and 4 month). Due to the irregular distribution of gaging stations, layers below the Surficial Aquifer display only a portion of the contours within the model area. Conclusively, the forecasted heads matched reasonable well with those of historical heads for l=1, but forecasting errors increased significantly for l>1.

6. Conclusions

This paper presents a new form the STARX model with its ML fitting method by the EMSSE algorithm for a complete data set. The proposed EMSSE algorithm can estimate both model parameters and error covariances simultaneously, so that this estimation algorithm can be easily extended to the Kalman filter forecasting. The EMSSE algorithm, compare to the time-variant version, enables not only to accelerate convergence of parameter estimation, but to reduce both computational load and memory requirement. Therefore, this algorithm can ideally suitable for the large scale problems.

The presented methodologies were applied to the multi-layered aquifer system in Collier County, Florida, with discussions on the identification, calibration, and verification processes of the STARX model. Verification of the fitted STARX models was focused on both the data generation and forecasting abilities. Results of verification showed that the fitted STARX models preserved the first four moments as well as space-time correlations of the historical data. Comparison of the forecasted and historical heads showed reasonable agreements particularly for the small lead time.

Appendix: Expectation-Maximization Algorithm with the Kalman Smoothed Estimator

With the state-space format of the STARX model in equations (6) and (7), the modified Kalman smoother estimator

$$\hat{x}_t^T = E\,[x(t) \mid y_1,\ldots,y_T,\ z_1,\ldots,z_T] \tag{A1}$$

is obtained from the minimizing the mean square error

$$p_t^T = E\,[(x_t^* - \hat{x}_t^T)(x_t^* - \hat{x}_t^T)' \mid y_1,\ldots,y_T,\ z_1,\ldots,z_T] \tag{A2}$$

and the error covariance for one-time lagging state vectors

$$p_{t,t-1}^T = E\,[(x_t^* - \hat{x}_t^T)(x_{t-1}^* - \hat{x}_{t-1}^T)' \mid y_1,\ldots,y_T,\ z_1,\ldots,z_T]. \tag{A3}$$

The detailed derivations of the above conditional expectations can be found from *Jazwinski* [1970, page 201] and *Shumway and Stoffer* [1981]. The final results applied to the STARX model are as follows: First, the forward recursion for t=1,...,T, is given by

$$\hat{x}_t^{t-1} = \Phi \hat{x}_{t-1}^{t-1} + \Psi z(t) \tag{A4}$$

$$p_t^{t-1} = \Phi p_{t-1}^{t-1} \Phi' + Q^* \tag{A5}$$

$$K_t = p_t^{t-1} M(t)' \,[M(t)\, p_t^{t-1} M(t)' + R]^{-1} \tag{A6}$$

$$\hat{x}_t^t = \hat{x}_t^{t-1} + K_t\,[y_t - M(t)\,\hat{x}_t^{t-1}] \tag{A7}$$

$$p_t^t = p_t^{t-1} + K_t\, M(t)\, p_t^{t-1} \tag{A8}$$

where $Q^* = E[w(t)\ w(t)'] = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$, $Q=E[w_t\ w_t']$, *and* $R=E[v_t\ v_t']$.

In order to calculate the smoothers, the backward recursion for t=T,T-1,...,1, is performed by

$$J_{t-1} = p_{t-1}^{t-1}\Phi'(p_t^{t-1})' \tag{A9}$$

$$\hat{x}_{t-1}^{T} = \hat{x}_{t-1}^{t-1} + J_{t-1}\ (\hat{x}_t^{T} - \Phi\hat{x}_t^{t-1}) \tag{A10}$$

$$p_{t-1}^{T} = p_{t-1}^{t-1} + J_{t-1}\ (p_t^{T} - p_t^{t-1})\ J_{t-1}' \tag{A11}$$

Also, an one-time lagged smoothed error covariances, for t=T,T-1,...,2, is

$$p_{t-1,t-2}^{T} = p_{t-1}^{t-1}J_{t-2}' + J_{t-1}[p_{t,t-1}^{T} - \Phi p_{t-1}^{t-1}]\ J_{t-2}' \tag{A12}$$

$$\textit{with,}\quad p_{T,T-1}^{T} = [I-K_T M(T)]\ \Phi\ p_{T-1}^{T-1} \tag{A13}$$

where, K_T and p_{T-1}^{T-1} are those from the last time step in the forward recursion. The equation (5) of the STARX model can be given by [see Stoffer, 1985]

$$\begin{aligned} Q(\theta|\theta_i) = {} & -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}tr\ \{\Sigma^{-1}[\,p^{T}(0) + (x^{T}(0)-\mu)\ (x^{T}(0)-\mu)']\} \\ & -\frac{T}{2}\ln|Q| - \frac{1}{2}tr\ \{Q^{-1}\ [C - \theta S_t(1)' - S_t(1)\theta' + \theta S_{t-1}(0)\theta']\} \\ & -\frac{T}{2}\ln|R| - \frac{1}{2}tr\ \{R^{-1}\sum_{t=1}^{T}[(y_t-M(t)\hat{x}_t^{T})(y_t-M(t)\hat{x}_t^{T})' + M(t)p_t^{T}M(t)']\} \end{aligned} \tag{A14}$$

where $x^{T}(0)\approx N(\mu,\Sigma)$, $|A|$ is the determinant of a matrix A, trA is the trace of A, and

$$S_{t-1}(0) = \begin{bmatrix} A & F \\ F' & H \end{bmatrix} \tag{A15}$$

$$S_t(1) = [B \; G] \tag{A16}$$

with $A=\sum_{t=1}^{T}(p_{t-1}^{T}+\hat{x}_{t-1}^{T}\hat{x}_{t-1}^{T'})$, $B=\sum_{t=1}^{T}(p_{t,t-1}^{T}+\hat{x}_{t,1}^{T}\hat{x}_{t-1}^{T'})$, $C=\sum_{t=1}^{T}(p_{t}^{T}+\hat{x}_{t,1}^{T}\hat{x}_{t,1}^{T'})$, $F=\sum_{t=1}^{T}[\hat{x}_{t-1}^{T}z(t)']$,

$G=\sum_{t=1}^{T}[\hat{x}_{t,1}^{T}z(t)']$, and $H=\sum_{t=1}^{T}[z(t)z(t)']$.

with $\hat{x}_{t,1}^{T}$ is the first sub-vector in the $\hat{x}_{t}^{T\prime}=[\hat{x}_{t,1}^{T\prime},\ldots,\hat{x}_{t,Nq+1}^{T\prime}]$. The dimensions are: $S_{t-1}(0)$ (na×na), $S_t(1)$ (nx×na), A(ns×), B (nx×ns), C (nx×nx), F (ns×nc), G (nx×nc), and H (nc×nc), with na=ns+nc. The maximization step is obtained by maximizing (A14) with respect to the parameters θ, Q, and R. The resulting regression estimators are:

$$\theta_{i+1} = D^{*} \circ [S_t(1) \; S_{t-1}(0)'] \tag{A17}$$

$$Q_{i+1} = \frac{1}{T}[C - \theta_{i+1}S_t(1)' - S_t(1)\theta_{i+1}' + \theta_{i+1}S_{t-1}(0)\theta_{i+1}'] \tag{A18}$$

$$R_{i+1} = \frac{1}{T}\sum_{t=1}^{T}\{[y_t - M(t)\hat{x}_t^{T}][(y_t - M(t)\hat{x}_t^{T}]' + M(t)\, p_t^{T}\, M(t)'\} \tag{A19}$$

where $D^{*} = \begin{bmatrix} D_1 & \cdots & D_{Nq-1} & D_{Nq} & E_0 & \cdots & E_{Nk} \\ I & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & 0 & \vdots & \ddots & \vdots \\ 0 & \cdots & I & 0 & 0 & \cdots & 0 \end{bmatrix}$.

References

Bellman, R, *Methods of Nonlinear Analysis,* Volume I, Academic Press, New York, 1970.

Bennett, R. J. *Spatial Time Series: Analysis-Forecasting-Control*, Pion Ltd., 207 Brondesbury Park, London, 1979.

Bennett M. W., A three-dimensional finite difference ground water flow model of western Collier county, Florida, Tech. Pub. #92-04, South Florida Water Management District, West Palm Beach, Florida, 1992.

Box, G. E. P., and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., Prentice-Hall, Englewood Cliffs, N.J., 1976.

Bras, R. L., and I. Rodriguez-Iturbe, *Random Functions and Hydrology*, Addison-Wesley, Reading, Mass., 1985.

Brockwell, P. J., and R. A. Davis, *Time Series Theory and Methods*, Springer-Verlag, N.Y., 1987.

Cressie, N. A. C., *Statistics for Spatial Data.* John Wiley & Sons Inc., New York, 1991.

Dempster, A. P., N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of Royal Statistical Society* B, 39, 1-38, 1977.

Environmental Systems Research Institute, "*Understanding GIS: The Arc/Info Method*", Environmental Systems Research Institute Inc., CA, 1992.

Graham, W.D., and C. Tankersley, Forecasting piezometric head levels in the Floridan aquifer: A Kalman filtering approach, *Water Resour. Res.*, 29(11), 3791-3800, 1993.

Gupta, N. K., and Mehra, R.K., Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculation, *IEEE Trans. on Automatic Contl.*, 19, 1970.

Harvey, A. C., *The Econometric Analysis of Time Series*, Second Edition, The MIT Press,

Cambridge, Massachusetts, 1990.

Horn, R.A., and C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, N.Y., 1985.

Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

Lainiotis, D. G., Partitioned Filters, *Applications of Kalman Filter to Hydrology, Hydraulics, and Water Resources*, p.71-81, edited by C. Chiu, Univ. of Pittsburgh, Pittsburgh, PA, 1978.

Pfeifer, P. E. and Deutsch, S. J. Identification and interpretation of first order space-time ARMA modelling, *Technometrics*, 22, 397-408, 1980.

Salas, J. D., J. W. Delleur, V. Yevjevich, and L. W. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Fort Collins, Colo., 1980.

Shih, G., W. S. Burns, and R. F. Bower, A ground water drought management model for Collier county, Florida, Tech. Pub. #92-01, South Florida Water Mgt. District, West Palm Beach Florida, 1992.

Shumway, R. H., *Applied Statistical Time Series Analysis*, Prentice Hall, Englewood Cliffs, N.J., 1988.

Shumway, R. H., and D. S. Stoffer, Time series smoothing and forecasting using the EM algorithm, Tech. Report No. 27, Div of Statistics, University of California, Davis, California, 1981.

Stoffer, D. S. Maximum likelihood fitting of STARMAX models to incomplete space-time series data, *Time Series Analysis: Theory and Practice* 6, page 283-296, edited by O. D. Anderson, J. K. Ord, and E. A. Robinson, North-Holland, Amsterdam, 1985.

Stoffer, D. S. Estimation and identification of space-time ARMAX models in the presence of missing data. *Journal of the American Statistical Association*, 81, 762-772, 1986.

List of Figures

Table 1. Approximate Hydraulic Properties in Each Aquifer

Layer	Name of aquifer	Aquifer thickness (meters)	K_H or T* (m^2/day)	# of selected gaging stations
Layer 1	Surficial aquifer	12-18	30-1000	48
Layer 2	Lower Tamiami aquifer	20-400	900-3000	43
Layer 3	Sandstone aquifer	46-76	15-2300	14
Layer 4	Mid-Hawthorn aquifer	50-125	46-110	10

* K_H is the horizontal hydraulic conductivity, and T is the transmissivity in the confined aquifer.

Table 2. Maximum likelihood values with information criteria.

Nq	Nk	The 1st term of -2lnL(A)	The 2nd term of -2lnL(B)	-2lnL (A)+(B)	AIC	BIC
			Model I, nx=48			
1	-	-13678	2977	-10701	-91.0	-22.4
2	**-**	**-15956**	**613**	**-15343**	**-97.9**	**39.3**
3	-	-15638	236	-15402	-41.1	164.7
4	-	-16288	122	-16166	5.9	280.3
			Model II, nx=43, nz=48			
1	-	-11667	3111	-8555	-72.6	-17.5
2	-	-16558	853	-15705	-125.6	-15.1
3	-	-17628	344	-17284	-101.4	63.8
1	**0**	**-17814**	**1420**	**-16394**	**-147.1**	**-51.0**
2	0	-15076	334	-14742	-77.9	73.2
3	0	-15792	140	-15652	-44.3	161.8
1	1	-16418	193	-16225	-110.3	26.7
2	1	-12188	62	-12125	-7.2	184.9
3	1	-14546	34	-14512	5.9	253.0
			Model III, nx=24, nz=43			
1	-	-3673	1871	-1802	-10.6	6.5
2	-	-5644	1802	-3843	-24.6	9.7
3	-	-8765	1274	-7491	-60.8	-9.4
1	0	-5350	1833	-3517	-23.6	6.4
2	0	-8406	1555	-6852	-55.6	-8.4
3	**0**	**-9716**	**600**	**-9116**	**-72.6**	**-8.3**
1	1	-7530	1753	-5777	-44.2	-1.4
2	1	-9336	662	-8674	-70.1	-10.0
3	1	-8941	307	-8634	-55.1	22.1
2	2	-8452	218	-8234	-53.2	19.7
3	2	-8779	99	8680	-45.0	45.1
3	3	-6753	15	-6738	-7.2	95.7

Ref.: The selected models are marked by the bold face.

Table 3. Spatial mean and standardization deviations of the standardized forecasted error, $p_0^{1/2}$.

Lead Time l	Model I		Model II		Model III	
	Mean	S.D.*	Mean	S.D.*	Mean	S.D.*
<1	0.260	0.044	0.251	0.062	0.242	0.066
1	0.421	0.067	0.640	0.232	0.641	0.339
2	0.484	0.071	0.651	0.232	0.684	0.319
3	0.526	0.075	0.652	0.232	0.720	0.292
4	0.544	0.078	0.652	0.232	0.734	0.281
5	0.559	0.081	0.652	0.232	0.738	0.279
6	0.570	0.084	0.652	0.232	0.742	0.276
7	0.579	0.087	0.652	0.232	0.743	0.275
8	0.586	0.092	0.652	0.232	0.744	0.275
9	0.592	0.096	0.652	0.232	0.744	0.274
10	0.596	0.099	0.652	0.232	0.744	0.274
11	0.601	0.103	0.652	0.232	0.744	0.274
12	0.604	0.106	0.652	0.232	0.744	0.274

* S.D. is the standard deviation

Figure 1

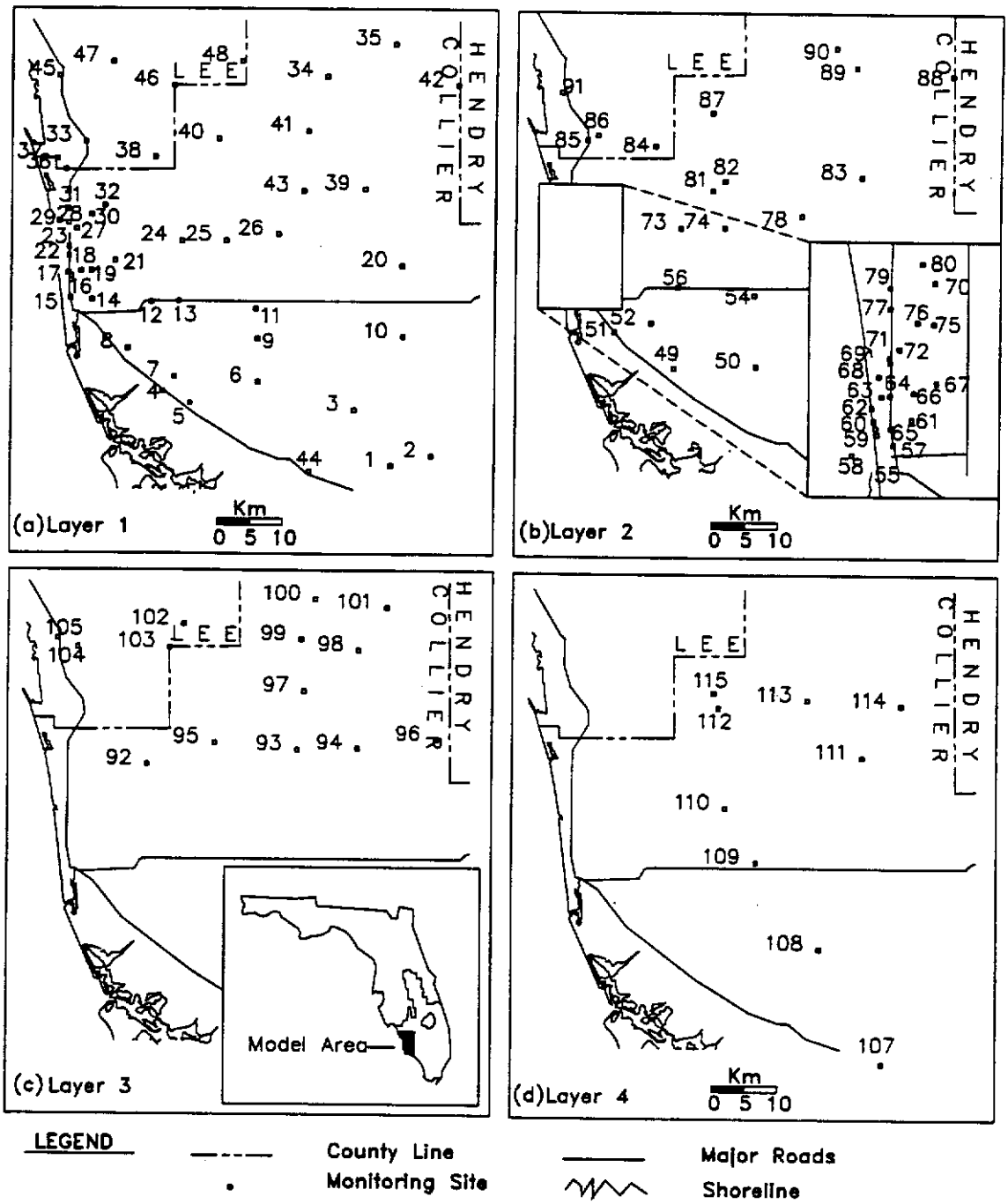

LEE
HENDRY
COLLIER
(a)Layer 1
(b)Layer 2
(c)Layer 3
(d)Layer 4
Km
0 5 10
Model Area
LEGEND
County Line
Monitoring Site
Major Roads
Shoreline

Figure 2a

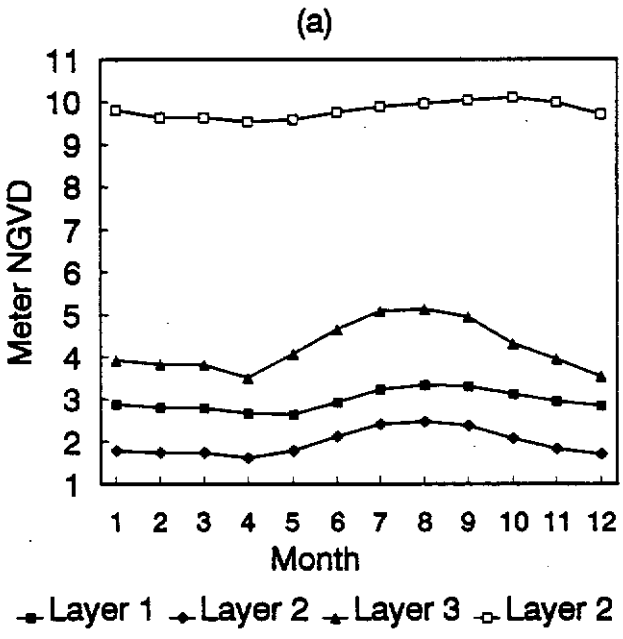

(a)
Meter NGVD
11
10
9
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10 11 12
Month
Layer 1
Layer 2
Layer 3
Layer 2

Figure 2b

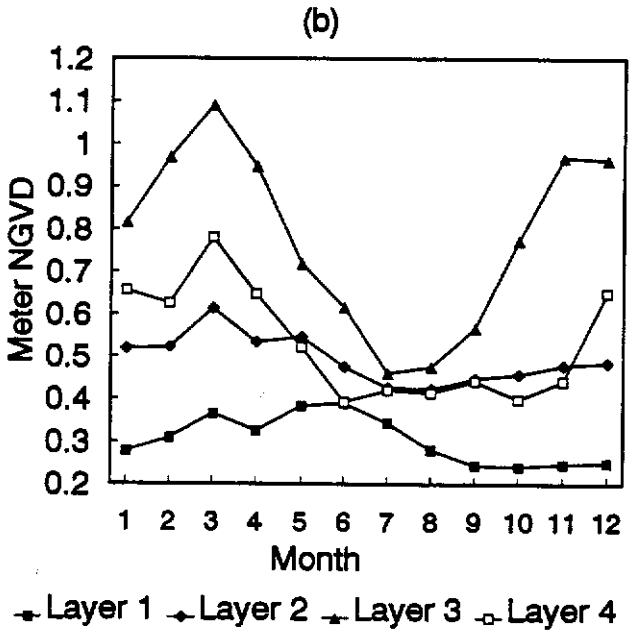

(b)
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Meter NGVD
1 2 3 4 5 6 7 8 9 10 11 12
Month
Layer 1
Layer 2
Layer 3
Layer 4

Figure 3

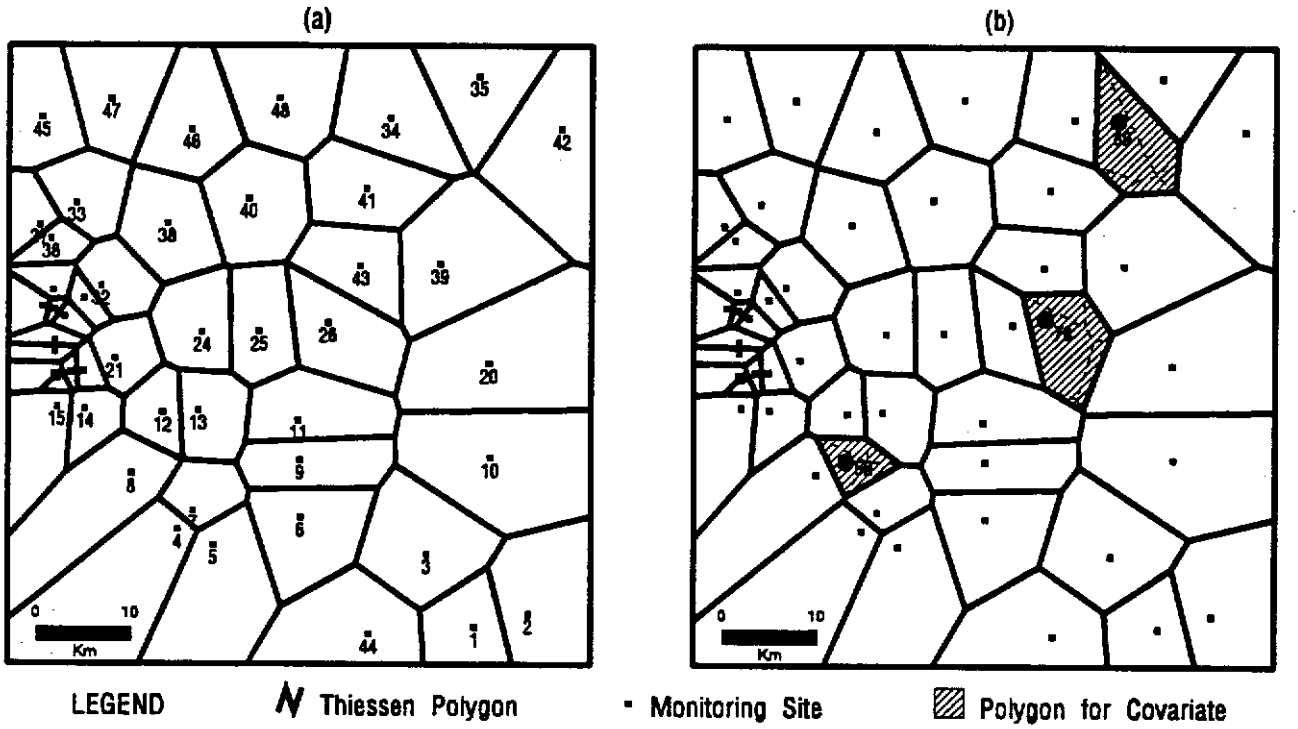

(a)
(b)
0
10
Km
LEGEND
Thiessen Polygon
Monitoring Site
Polygon for Covariate

Figure 4

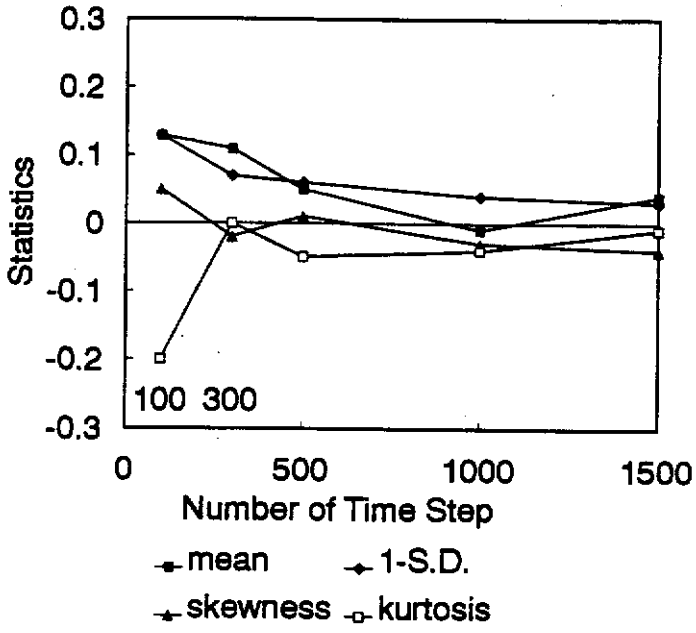

0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Statistics
100 300
0
500
1000
1500
Number of Time Step
mean
1-S.D.
skewness
kurtosis

Figure 5a

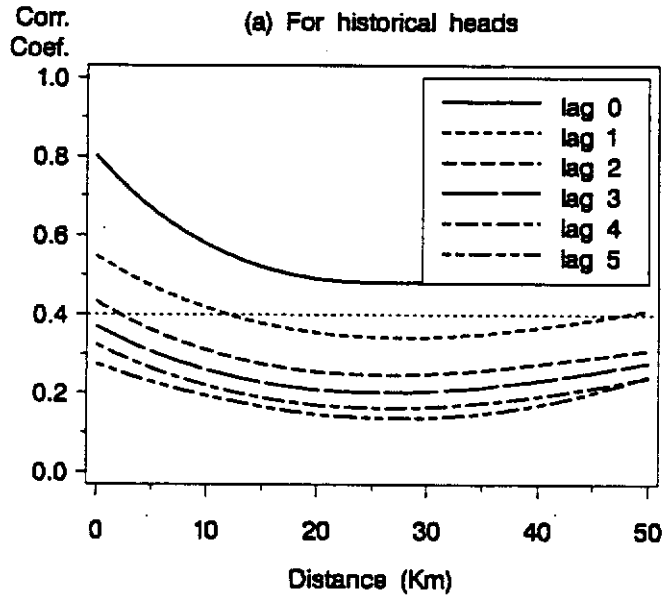

(a) For historical heads
Corr. Coef.
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Distance (Km)
lag 0
lag 1
lag 2
lag 3
lag 4
lag 5

Figure 5b

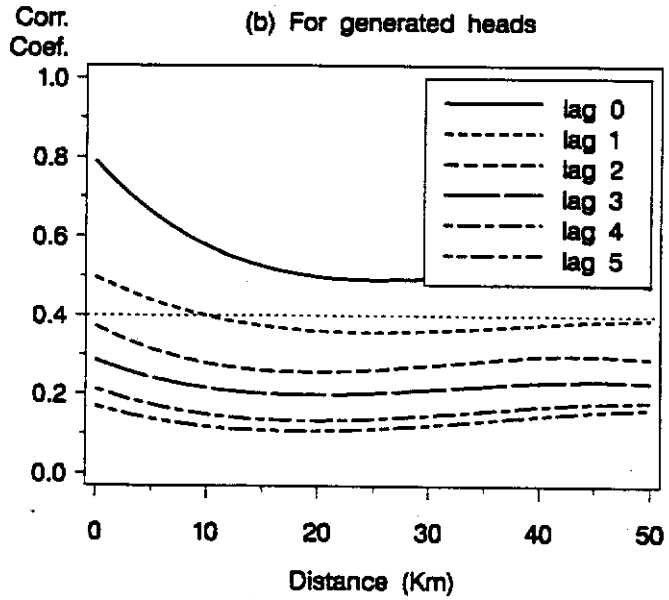
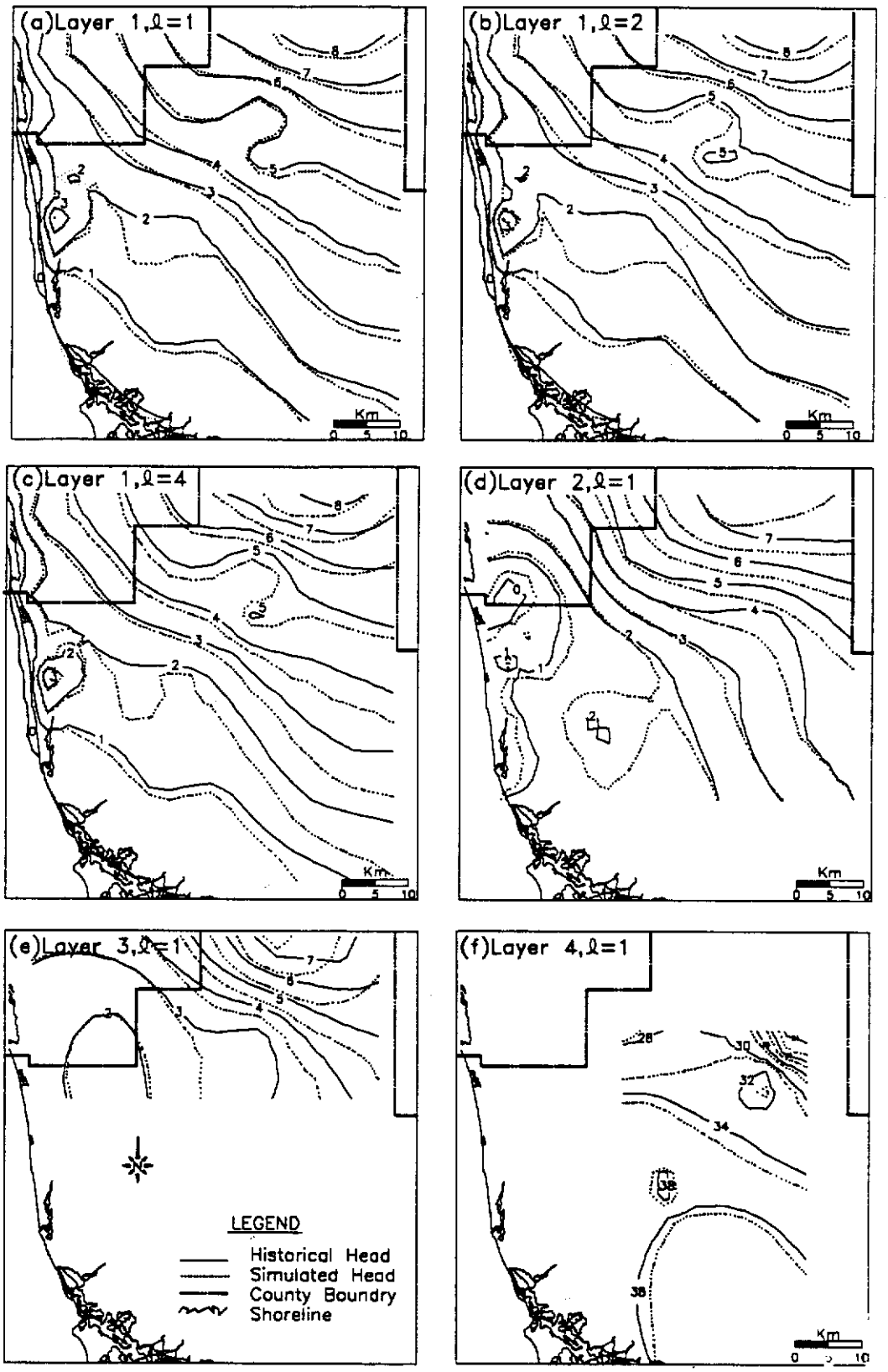

Corr.
Coef.
(b) For generated heads
1.0
0.8
0.6
0.4
0.2
0.0
lag 0
lag 1
lag 2
lag 3
lag 4
lag 5
0
10
20
30
40
50
Distance (Km)

Figure 6


(a)Layer 1,ℓ=1
(b)Layer 1,ℓ=2
(c)Layer 1,ℓ=4
(d)Layer 2,ℓ=1
(e)Layer 3,ℓ=1
(f)Layer 4,ℓ=1
LEGEND
Historical Head
Simulated Head
County Boundry
Shoreline
Km
0 5 10